

RESEARCH

Open Access



Convergence rate for the moving least-squares learning with dependent sampling

Qin Guo^{1*} and Peixin Ye¹

*Correspondence:

guoqin_1985@163.com;
guoqin@mail.nankai.edu.cn

¹School of Mathematical Sciences
and LPMC, Nankai University,
Tianjin, China

Abstract

We consider the moving least-squares (MLS) method by the regression learning framework under the assumption that the sampling process satisfies the α -mixing condition. We conduct the rigorous error analysis by using the probability inequalities for the dependent samples in the error estimates. When the dependent samples satisfy an exponential α -mixing, we derive the satisfactory learning rate and error bound of the algorithm.

MSC: 68T05; 68Q32; 60E15

Keywords: Moving least-squares; Regression function; Mixing sequence; Probability inequality; Error bound

1 Introduction

The least-squares (LS) method is an important global approximate method based on the regular or concentrated data sample points. However, there are still some irregular or scattered samples which are obtained in many practical applications such as engineering and machine learning [1–4]. They also need to be analyzed to achieve their special usefulness. The moving least-squares (MLS) method was introduced by McLain in [4] to draw a set of contours based on a cluster of scattered data sample points. It turns out that the MLS method is a useful local approximation tool in various fields of mathematics such as approximation theory, data smoothing [5], statistics [6], and numerical analysis [7]. Recently, research effort has been made to study the regression learning algorithm by the MLS method, see [8–12]. The main advantage of the MLS regression learning algorithm is that we can learn the regression function in the simple function space, usually generated by polynomials.

We recall the regression learning problem by the MLS method briefly. Functions for learning are defined on a compact metric space X (input space) and take values in $Y = \mathbb{R}$ (output space). The sampling process is controlled by an unknown Borel probability measure ρ on $Z = X \times Y$. We define the regression function as follows:

$$f_{\rho}(x) = \int_Y y d\rho(y|x),$$

where $\rho(\cdot|x)$ is the conditional probability measure induced by ρ on Y given $x \in X$. The goal of regression learning is to find a good approximation of the regression function f_ρ based on a set of random samples $\mathbf{z} = \{z_i\}_{i=1}^m = \{(x_i, y_i)\}_{i=1}^m \in Z^m$ drawn according to the measure ρ .

We define the approximation f_z of f_ρ pointwisely:

$$f_z(x) := f_{z, \sigma, x}(x) = \arg \min_{f \in \mathcal{H}} \mathcal{E}_{z, x}(f), \quad x \in X, \tag{1.1}$$

the local moving empirical error is defined by

$$\mathcal{E}_{z, x}(f) = \frac{1}{m} \sum_{i=1}^m \Phi\left(\frac{x}{\sigma}, \frac{x_i}{\sigma}\right) (f(x_i) - y_i)^2, \tag{1.2}$$

where the hypothesis space $\mathcal{H} \subseteq C(X)$ is a \tilde{d} -dimensional Lipschitz function space, $\sigma = \sigma(m) > 0$ is a window width, and $\Phi : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^+$ is called an MLS weight function which satisfies the conditions as follows, see [9, 10]:

$$(1) \quad \int_{\mathbb{R}^n} \Phi(x, t) dt = 1, \quad \forall x, t \in \mathbb{R}^n, \tag{1.3}$$

$$(2) \quad \Phi(x, t) \geq c_q, \quad \forall |x - t| \leq 1, \tag{1.4}$$

$$(3) \quad |\Phi(x, t)| \leq \frac{\tilde{c}_q}{(1 + |x - t|)^q}, \quad \forall x, t \in \mathbb{R}^n, \tag{1.5}$$

where the constants $q > n + 1, c_q, \tilde{c}_q > 0$.

The task of the paper is to derive the error bound of $\|f_z - f_\rho\|_{\rho_X}$ with the norm $\|f(\cdot)\|_{\rho_X} := (\int_X |f(\cdot)|^2 d\rho_X)^{\frac{1}{2}}$ to evaluate the approximation ability of f_z , see [13–22]. The error analysis of algorithm (1.1) for the independent and identical (i.i.d.) samples has been carried out in [8–10]. However, the samples are not independent but are not far from being independent in some real data analysis such as market prediction, system diagnosis, and speech recognition. The mixing conditions can quantify how close to independence a sequence of random samples is. In [14, 16, 23–25], the authors carried out the regression estimation of the least squares algorithm with the α -mixing samples. Up to now there has been no result of algorithm (1.1) obtained in the case of dependent samples. Hence we extend the analysis of algorithm (1.1) to the α -mixing sampling setting which is quite easy to establish, see [26].

Definition 1.1 Let \mathcal{M}_a^b denote the σ -algebras of events generated by the random samples $\{z_i = (x_i, y_i)\}_{i=a}^b, \{z_i\}_{i \geq 1}$ is said to satisfy a strongly mixing condition (or α -mixing condition) if

$$\alpha_i = \sup_{k \geq 1} \sup_{A \in \mathcal{M}_1^k, B \in \mathcal{M}_{k+i}^\infty} |P(A \cap B) - P(A)P(B)| \longrightarrow 0, \quad \text{as } i \rightarrow \infty. \tag{1.6}$$

Specifically, if there exist some positive constants $\bar{\alpha} > 0, \beta > 0$, and $c > 0$ such that

$$\alpha_i \leq \bar{\alpha} \exp(-ci^\beta), \quad \forall i \geq 1, \tag{1.7}$$

then it is said to satisfy an exponential strongly mixing condition.

Our goal is to obtain the convergence rate as $m \rightarrow \infty$ of algorithm (1.1) under hypothesis (1.7). The rest of the paper is organized as follows. In Sect. 2, we review some concepts and state our main results and the error decomposition. In Sect. 3, we present the estimate of the sample error. In Sect. 4, we provide the proofs of the main results.

2 Main results and error decomposition

Before giving the main results, we firstly need to provide some concepts that will be referred to throughout this paper, see [8–10].

Definition 2.1 The probability measure ρ_X on X is said to satisfy the condition L_τ with exponent $\tau > 0$ if

$$\rho_X(B(x, r)) \geq c_\tau r^\tau, \quad \forall 0 < r \leq r_0, x \in X, \tag{2.1}$$

where the constants $r_0 > 0, c_\tau > 0$, and $B(x, r) = \{u \in X : |u - x| \leq r, \text{ for } r > 0\}$.

Definition 2.2 We say that the hypothesis space \mathcal{H} satisfies the norming condition with exponent $\zeta > 0$ and $d \in \mathbb{N}$ if we can find points $\{u_i\}_{i=1}^d \subset B(x, \sigma)$ for every $x \in X$ and $0 < \sigma \leq \sigma_0$ satisfying $|u_i - u_j| \geq 2c_{\mathcal{H}}\sigma$ for $i \neq j$ and

$$\left(\sum_{i=1}^d |f(u_i)|^2 \right)^{\frac{1}{2}} \geq c_{\mathcal{H}} \sigma^\zeta \|f\|_{C(X)}, \quad \forall f \in \mathcal{H}, \tag{2.2}$$

where the constants $\sigma_0 > 0, c_{\mathcal{H}} > 0$ and d is chosen as at least the dimension \tilde{d} of \mathcal{H} .

Here we assume $|y| \leq M$ almost surely, and all the constants such as $\tilde{C}, C_{\mathcal{H}, \zeta}, A_{\tau, \zeta}, C_{\mathcal{H}, \rho_X}, C'_{\mathcal{H}, \rho_X}$, and so on are independent of the key parameters δ, m , or σ in this paper. Now we give our main results of algorithm (1.1).

Theorem 2.1 Assume that (1.7), (2.1), and (2.2) hold. Suppose $0 < p < 2, \sigma = (m^{(\alpha)})^{-\gamma}$ with $m^{(\alpha)} = \lfloor m \lceil \{\frac{8m}{c}\}^{1/(1+\beta)} \rceil^{-1} \rfloor, \gamma > 0$, and $0 < \sigma \leq \min\{\sigma_0, 1, (r_0/C_{\mathcal{H}, \zeta})^{1/\max\{\zeta, 1\}}\}$. If m satisfies

$$\begin{aligned} (m^{(\alpha)})^{1-\gamma \tau \max\{\zeta, 1\}} &\geq (256c_p^{\frac{1}{p}}/3 + A_{\tau, \zeta})(\log(2 + 8e^{-2\bar{\alpha}})/\delta)^{1+\frac{1}{p}} \\ &\quad + A_{\tau, \zeta} \gamma \log m^{(\alpha)}, \end{aligned} \tag{2.3}$$

then for any $0 < \delta < 1$, with confidence $1 - \delta$, we have

$$\|f_z - f_\rho\|_{\rho_X}^2 \leq \tilde{C} (m^{(\alpha)})^{4\gamma(\zeta + \frac{5}{2} \max\{\zeta, 1\}) - n\gamma - \frac{1}{p+1}}. \tag{2.4}$$

Then we can obtain the explicit learning rate of algorithm (1.1) with selecting the suitable parameter $\sigma = \sigma(m)$.

Theorem 2.2 Under the assumptions of Theorem 2.1, if we choose $\sigma = (m^{(\alpha)})^{-\frac{\varepsilon}{(4\zeta + 2\max\{\tau, \zeta\})}}$, $0 < \varepsilon < 1/4$, and

$$m^{(\alpha)} \geq C_1 \left((\log(2 + 8e^{-2\bar{\alpha}})/\delta)^{1+\frac{1}{p}} + \log m^{(\alpha)} \right)^2 + \sigma_0^{-(4\zeta + 2\max\{\tau, \zeta\})/\varepsilon}, \tag{2.5}$$

then with confidence $1 - \delta$, we have

$$\|f_{\mathbf{z}} - f_{\rho}\|_{\rho_X} \leq C_2(m^{(\alpha)})^{\varepsilon - \frac{1}{2}}, \tag{2.6}$$

where

$$C_1 = \left(A_{\tau, \zeta} + \frac{256}{3} c_p^{-\frac{1}{p}} \right)^2 \left(1 + \frac{1}{4\zeta + 2 \max\{\tau, \tau\zeta\}} \right)^2. \tag{2.7}$$

Remark 2.1 The result of the above theorem shows that the learning rate tends to $m^{-\frac{1}{2}}$ when $\sigma \rightarrow 1$. For the i.i.d. case, the same rate has been obtained in [9, 10].

To estimate the quantity of the total error $\|f_{\mathbf{z}} - f_{\rho}\|_{\rho_X}$, we use the proposition from [8] below.

Proposition 2.1 *Assume (2.1) and (2.2) hold. Then we have*

$$\|f_{\mathbf{z}} - f_{\mathcal{H}}\|_{\rho_X}^2 \leq \widetilde{C}_{\mathcal{H}} \sigma^{-2\zeta - \tau \max\{\zeta, 1\}} \int_X (\mathcal{E}_x(f_{\mathbf{z}, \sigma, x}) - \mathcal{E}_x(f_{\mathcal{H}, \sigma, x})) d\rho_X(x), \tag{2.8}$$

where

$$\mathcal{E}_x(f) = \int_Z \Phi\left(\frac{x}{\sigma}, \frac{u}{\sigma}\right) (f(u) - y)^2 d\rho(u, y), \quad \forall f : X \rightarrow \mathbb{R} \tag{2.9}$$

is called the local moving expected risk and

$$f_{\mathcal{H}}(x) := f_{\mathcal{H}, \sigma, x} = \arg \min_{f \in \mathcal{H}} \mathcal{E}_x(f), \quad x \in X, \tag{2.10}$$

is called the target function.

Remark 2.2 Here we assume $f_{\rho} \in \mathcal{H}$. It follows from

$$\mathcal{E}_x(f) = \int_X \Phi\left(\frac{x}{\sigma}, \frac{u}{\sigma}\right) (f(u) - f_{\rho}(u))^2 d\rho_X(u) + \mathcal{E}_x(f_{\rho}), \quad \forall f : X \rightarrow \mathbb{R}, \tag{2.11}$$

that $f_{\mathcal{H}} = f_{\rho}$. Thus $\|f_{\mathbf{z}} - f_{\rho}\|_{\rho_X} = \|f_{\mathbf{z}} - f_{\mathcal{H}}\|_{\rho_X}$.

Next we only need to provide the upper bound of the integral in (2.8). So to do this, we give its decomposition as follows:

$$\begin{aligned} \int_X (\mathcal{E}_x(f_{\mathbf{z}, \sigma, x}) - \mathcal{E}_x(f_{\mathcal{H}, \sigma, x})) d\rho_X(x) &\leq \int_X [(\mathcal{E}_x(f_{\mathbf{z}, \sigma, x}) - \mathcal{E}_x(f_{\mathcal{H}, \sigma, x})) \\ &\quad - (\mathcal{E}_{\mathbf{z}, x}(f_{\mathbf{z}, \sigma, x}) - \mathcal{E}_{\mathbf{z}, x}(f_{\mathcal{H}, \sigma, x}))] d\rho_X(x) \\ &:= \mathcal{S}(\mathbf{z}, \sigma). \end{aligned} \tag{2.12}$$

What is left is to estimate the sample error $\mathcal{S}(\mathbf{z}, \sigma)$.

3 Estimates for the sample error

In order to obtain the probability estimate of $\mathcal{S}(\mathbf{z}, \sigma)$, we shall use the upper bound for $f_{\mathbf{z}, \sigma, x}$ and $f_{\mathcal{H}, \sigma, x}$. We firstly derive the confidence-based estimate of $f_{\mathbf{z}, \sigma, x}$ as follows.

Proposition 3.1 *Under the assumptions of Theorem 2.1, if*

$$m^{(\alpha)} \geq -A_{\tau, \zeta} \log(\delta \sigma) \sigma^{-\tau \max\{\zeta, 1\}}, \tag{3.1}$$

then with confidence at least $1 - \delta$, we have

$$\|f_{\mathbf{z}, \sigma, x}\|_{C(X)} \leq \frac{2^{3+\tau+\zeta} M}{\sqrt{c_\tau c_q} C_{\mathcal{H}, \zeta}^{\tau/2} c_{\mathcal{H}}} \sigma^{-\zeta - \max\{\frac{\tau}{2}, \frac{\tau \zeta}{2}\}} := C_{\mathcal{H}, \rho_X} \sigma^{-\zeta - \max\{\frac{\tau}{2}, \frac{\tau \zeta}{2}\}}, \quad \forall x \in X, \tag{3.2}$$

where

$$C_{\mathcal{H}, \zeta} = \min \left\{ \frac{c_{\mathcal{H}}}{2^{\zeta+1} \sqrt{d} C_{\mathcal{H}, 0}}, \frac{c_{\mathcal{H}}}{2}, \frac{1}{2} \right\},$$

$$A_{\tau, \zeta} = 2^{\tau+1} (c_\tau C_{\mathcal{H}, \zeta}^\tau - 2^{\tau+1})^{-1} \left[\frac{7}{6} + \frac{7}{6} \log(1 + 4e^{-2\bar{\alpha}}) + \frac{7n}{6} \log \left(1 + \frac{4B_X}{C_{\mathcal{H}, \zeta}} \right) \right].$$

The proof is analogous to that of Theorem 3 in [8] except that we need to use the following Lemma 3.1 for the dependent sampling setting to replace Lemma 2 in [8].

Lemma 3.1 *Let $0 < r \leq r_0$ and $0 < \delta < 1$. If (1.7) and (2.1) hold, then with confidence $1 - \delta$, we have*

$$\frac{\sharp(\mathbf{x} \cap B(x, r))}{m} \geq c_\tau \left(\frac{r}{2} \right)^\tau + \frac{7 \log \delta - 7 \log(1 + 4e^{-2\bar{\alpha}}) - 7n \log(\frac{4B_X}{r} + 1)}{6m^{(\alpha)}} - 1, \quad \forall x \in X. \tag{3.3}$$

Specifically, if

$$m^{(\alpha)} > \frac{[7 \log(\frac{1}{\delta}) + 7 \log(1 + 4e^{-2\bar{\alpha}}) + 7n \log(\frac{4B_X}{r} + 1)]}{6(\frac{c_\tau r^\tau}{2^{\tau+1}} - 1)}, \tag{3.4}$$

then with confidence at least $1 - \delta$, we have

$$\frac{\sharp(\mathbf{x} \cap B(x, r))}{m} \geq \frac{c_\tau}{2^{\tau+1}} r^\tau, \quad \forall x \in X, \tag{3.5}$$

where $\frac{\sharp(\mathbf{x} \cap B(x, r))}{m}$ is the proportion of those sampling points lying in $B(x, r)$.

Proof It is shown in Theorem 5.3 of [27] that one can find $\{v_j\}_{j=1}^{\mathcal{N}} \subseteq X$ satisfying $X \subseteq B_R(\mathbb{R}^n) \subseteq \bigcup_{j=1}^{\mathcal{N}} B(v_j, \frac{r}{2})$ and $\mathcal{N} \leq (\frac{4R}{r} + 1)^n$. Let $\xi^{(j)} : X \rightarrow \mathbb{R}$ be the characteristic function of the set $B(v_j, \frac{r}{2})$. Its mean $\mu^{(j)} = \int_X \xi^{(j)}(x) d\rho_X = \rho_X(B(v_j, \frac{r}{2}))$ satisfies $|\xi^{(j)} - \mu^{(j)}| \leq 1$ and $\sigma^2(\xi^{(j)}) \leq 1$. Now we use the Bernstein inequality for the dependent samples in [28].

Proposition 3.2 *Suppose that (1.7) holds. Let the random variable $m^{(\alpha)}$ be the effective number of observations and $\xi_i = \xi(z_i)$ be a real-valued function on the probability space Z*

with mean $\mu = \int_Z \xi(z) d\rho$ and variance σ^2 . Assume that $|\xi_i - \mu| \leq D$ almost surely. Then, for every $\varepsilon > 0$,

$$P \left\{ \frac{1}{m} \sum_{i=1}^m [\xi_i - \mu] > \varepsilon \right\} \leq (1 + 4e^{-2\bar{\alpha}}) \exp \left\{ -\frac{m^{(\alpha)} \varepsilon^2}{2(\sigma^2 + \frac{1}{3}D\varepsilon)} \right\}. \tag{3.6}$$

Then it follows from the above proposition that

$$P \left\{ \frac{1}{m} \sum_{i=1}^m [\xi_i^{(j)} - \mu^{(j)}] \leq -\varepsilon \right\} \leq (1 + 4e^{-2\bar{\alpha}}) \exp \left\{ -\frac{m^{(\alpha)} \varepsilon^2}{2 + \frac{2}{3}\varepsilon} \right\}, \quad \forall \varepsilon > 0, \tag{3.7}$$

hence,

$$P \left\{ \min_{1 \leq j \leq \mathcal{N}} \left\{ \frac{1}{m} \sum_{i=1}^m [\xi_i^{(j)} - \mu^{(j)}] \right\} \leq -\varepsilon \right\} \leq \mathcal{N}(1 + 4e^{-2\bar{\alpha}}) \exp \left\{ -\frac{m^{(\alpha)} \varepsilon^2}{2 + \frac{2}{3}\varepsilon} \right\}. \tag{3.8}$$

For $0 < \delta < 1$, let

$$\mathcal{N}(1 + 4e^{-2\bar{\alpha}}) \exp \left\{ -\frac{m^{(\alpha)} \varepsilon^2}{2 + \frac{2}{3}\varepsilon} \right\} = \delta. \tag{3.9}$$

Then we get

$$\begin{aligned} \varepsilon &= \frac{\frac{2}{3} \log \frac{\mathcal{N}(1+4e^{-2\bar{\alpha}})}{\delta} + \sqrt{(\frac{2}{3} \log \frac{\mathcal{N}(1+4e^{-2\bar{\alpha}})}{\delta})^2 + 8m^{(\alpha)} \log \frac{\mathcal{N}(1+4e^{-2\bar{\alpha}})}{\delta}}}{2m^{(\alpha)}} \\ &\leq \frac{2 \log \frac{\mathcal{N}(1+4e^{-2\bar{\alpha}})}{\delta}}{3m^{(\alpha)}} + \sqrt{\frac{2 \log \frac{\mathcal{N}(1+4e^{-2\bar{\alpha}})}{\delta}}{m^{(\alpha)}}} \\ &\leq \frac{7 \log \frac{\mathcal{N}(1+4e^{-2\bar{\alpha}})}{\delta}}{6m^{(\alpha)}} + 1. \end{aligned} \tag{3.10}$$

It follows that, with confidence at least $1 - \delta$,

$$\min_{1 \leq j \leq \mathcal{N}} \left\{ \frac{1}{m} \sum_{i=1}^m [\xi_i^{(j)} - \mu^{(j)}] \right\} > -\frac{7 \log \frac{\mathcal{N}(1+4e^{-2\bar{\alpha}})}{\delta}}{6m^{(\alpha)}} - 1. \tag{3.11}$$

Hence, we have

$$\frac{1}{m} \sum_{i=1}^m [\xi_i^{(j)} - \mu^{(j)}] > -\frac{7 \log \frac{\mathcal{N}(1+4e^{-2\bar{\alpha}})}{\delta}}{6m^{(\alpha)}} - 1, \quad \forall j = 1, \dots, \mathcal{N}. \tag{3.12}$$

Condition (2.1) yields $\mu^{(j)} \geq c_\tau (\frac{r}{2})^\tau$. Also $\xi^{(j)}(x_i) = 1$ if $x_i \in B(v_j, \frac{r}{2})$ and 0 otherwise. So that $\frac{1}{m} \sum_{i=1}^m \xi^{(j)}(x_i) = \sharp(\mathbf{x} \cap B(v_j, \frac{r}{2}))/m$. Hence,

$$\sharp \left(\mathbf{x} \cap B \left(v_j, \frac{r}{2} \right) \right) / m > c_\tau \left(\frac{r}{2} \right)^\tau - \frac{7 \log \frac{\mathcal{N}(1+4e^{-2\bar{\alpha}})}{\delta}}{6m^{(\alpha)}} - 1, \quad \forall j = 1, \dots, \mathcal{N}. \tag{3.13}$$

Observe from $X \subseteq \bigcup_{j=1}^{\mathcal{N}} B(v_j, \frac{r}{2})$ that for each $x \in X$, there exists some $j \in 1, \dots, \mathcal{N}$ such that $x \in B(v_j, \frac{r}{2})$, i.e., $|v_j - x| \leq \frac{r}{2}$. Since $x_i \in B(v_j, \frac{r}{2})$ implies $|x_i - x| \leq |x_i - v_j| + |v_j - x| \leq r$, we see that

$$\begin{aligned} \#(\mathbf{x} \cap B(x, r)) / m &\geq \#(\mathbf{x} \cap B(v_j, \frac{r}{2})) / m \\ &\geq c_\tau \left(\frac{r}{2}\right)^\tau - \frac{7 \log \frac{\mathcal{N}(1+4e^{-2\alpha})}{\delta}}{6m^{(\alpha)}} - 1. \end{aligned} \tag{3.14}$$

This proves Lemma 3.1. □

Now we are in a position to prove Proposition 3.1.

Proof of Proposition 3.1 By (3.1) and setting $r = C_{\mathcal{H}, \zeta} \sigma^{\max\{\zeta, 1\}} \leq r_0$, it is easy to see that (3.4) holds. Then (3.5) is valid.

It follows from (3.5) and Definition 2.2 with σ replaced by $\frac{\sigma}{2}$ that

$$m_i / m = \#(\mathbf{x} \cap B(u_i, r)) / m > c_\tau r^\tau / 2^{\tau+1}, \tag{3.15}$$

and

$$|x_{i,l} - u_i| \leq r, \tag{3.16}$$

where $\{x_{i,l}\}_{l=1}^{m_i}$ are the points of the set $\#(\mathbf{x} \cap B(u_i, r))$, which implies

$$|x_{i,l} - x| \leq |x_{i,l} - u_i| + |u_i - x| \leq r + \frac{\sigma}{2} \leq \sigma, \tag{3.17}$$

where $x \in X$, $l = 1, \dots, \tilde{m}$, and $\tilde{m} = \min_{1 \leq i \leq d} \{m_i\}$.

Then, by (1.4), we have

$$\left| \Phi\left(\frac{x}{\sigma}, \frac{x_{i,l}}{\sigma}\right) \right| \geq c_q. \tag{3.18}$$

Hence

$$\begin{aligned} \frac{1}{m} \sum_{i=1}^m \Phi\left(\frac{x}{\sigma}, \frac{x_j}{\sigma}\right) (f_{z, \sigma, x}(x_j))^2 &\geq \frac{1}{m} \sum_{i=1}^d \sum_{l=1}^{\tilde{m}} \Phi\left(\frac{x}{\sigma}, \frac{x_{i,l}}{\sigma}\right) (f_{z, \sigma, x}(x_{i,l}))^2 \\ &\geq \frac{1}{m} \sum_{i=1}^d \sum_{l=1}^{\tilde{m}} c_q (f_{z, \sigma, x}(x_{i,l}))^2 \\ &\geq \frac{c_\tau}{2^{\tau+1}} r^\tau c_q \left(\frac{c_{\mathcal{H}} \sigma^\zeta}{2^{\zeta+1}}\right)^2 \|f_{z, \sigma, x}\|_{C(X)}^2. \end{aligned} \tag{3.19}$$

The last inequality has been proved in Theorem 3 in [8].

Finally, combining (3.19) with the following inequality

$$\frac{1}{m} \sum_{i=1}^m \Phi\left(\frac{x}{\sigma}, \frac{x_i}{\sigma}\right) (f_{z, \sigma, x}(x_i))^2 \leq \frac{2}{m} \sum_{i=1}^m \Phi\left(\frac{x}{\sigma}, \frac{x_i}{\sigma}\right) \{(0 - y_i)^2 + y_i^2\} \leq 4M^2, \tag{3.20}$$

we derive the desired result. □

We also need to invoke Lemma 4 in [8] which provides the result about the upper bound of $f_{\mathcal{H},\sigma,x}$.

Proposition 3.3 *Assume that (2.1) and (2.2) hold. Then, for some constant $C'_{\mathcal{H},\rho_X}$ independent of σ , we have*

$$\|f_{\mathcal{H},\sigma,x}\|_{C(X)} \leq C'_{\mathcal{H},\rho_X} \sigma^{-\zeta - \max\{\frac{\zeta}{2}, \frac{\zeta\zeta}{2}\}}, \quad \forall x \in X, 0 < \sigma \leq \min\{\sigma_0, 1\}. \tag{3.21}$$

Next we will bound the sample error. The estimation for $\mathcal{S}(\mathbf{z}, \sigma)$ relies on the ratio probability inequality below that can be found in [27].

Proposition 3.4 *Suppose that (1.7) holds. Let \mathcal{G} be a set of functions on Z and $c > 0$ such that, for each $g \in \mathcal{G}$, $\mu(g) = \int_Z g(z) d\rho \geq 0$, $\mu(g^2) \leq c\mu(g)$, and $|g(z) - \mu(g)| \leq D$ almost surely. Then, for every $\varepsilon > 0$ and $0 < \alpha \leq 1$, we have*

$$P\left\{ \sup_{g \in \mathcal{G}} \frac{\mu(g) - \frac{1}{m} \sum_{i=1}^m g(z_i)}{\sqrt{\varepsilon + \mu(g)}} \geq 4\alpha\sqrt{\varepsilon} \right\} \leq (1 + 4e^{-2\bar{\alpha}})\mathcal{N}(\mathcal{G}, \alpha\varepsilon) \times \exp\left\{ -\frac{\alpha^2 m^{(\alpha)} \varepsilon}{2c + \frac{2}{3}D} \right\}. \tag{3.22}$$

We obtain the upper bound estimate for $\mathcal{S}(\mathbf{z}, \sigma)$ by using Proposition 3.4.

Proposition 3.5 *If the assumptions of Proposition 3.1 hold,*

$$R = \max\{C_{\mathcal{H},\rho_X}, C'_{\mathcal{H},\rho_X}, M\} \sigma^{-\zeta - \frac{\zeta}{2} \max\{1, \zeta\}}, \tag{3.23}$$

and

$$m^{(\alpha)} \geq \frac{256}{3} c_p^{-\frac{1}{p}} \left(\log \frac{2 + 8e^{-2\bar{\alpha}}}{\delta} \right)^{1 + \frac{1}{p}}, \tag{3.24}$$

then with confidence $1 - \delta$, there holds

$$\int_X [(\mathcal{E}_x(f_{\mathbf{z},\sigma,x}) - \mathcal{E}_x(f_{\mathcal{H},\sigma,x})) - (\mathcal{E}_{\mathbf{z},x}(f_{\mathbf{z},\sigma,x}) - \mathcal{E}_{\mathbf{z},x}(f_{\mathcal{H},\sigma,x}))] d\rho_X(x) \leq 16R^2 D \sigma^n \left(\frac{256c_p}{3m^{(\alpha)}} \right)^{1/(1+p)} + \frac{1}{2} \int_X (\mathcal{E}_x(f_{\mathbf{z},\sigma,x}) - \mathcal{E}_x(f_{\mathcal{H},\sigma,x})) d\rho_X(x). \tag{3.25}$$

Proof Let the function $g(u, y)$ be defined on the function set

$$\mathcal{G}_R = \left\{ \int_X \Phi\left(\frac{x}{\sigma}, \frac{u}{\sigma}\right) [(f(u) - y)^2 - (f_{\mathcal{H},\sigma,x}(u) - y)^2] d\rho_X(x) : f \in B_R := \{f \in \mathcal{H} : \|f\|_{C(X)} \leq R\} \right\}. \tag{3.26}$$

With condition (1.5) and the bound c_ρ of the density function of ρ_X , we have

$$\int_X \Phi\left(\frac{x}{\sigma}, \frac{u}{\sigma}\right) d\rho_X(x) \leq c_\rho \tilde{c}_q \int_{\mathbb{R}^n} \frac{\sigma^n}{(1 + |u|)^q} du \leq \frac{2\pi^{n/2} c_\rho \tilde{c}_q}{(q - n)\Gamma(\frac{n}{2})} \sigma^n := D\sigma^n, \tag{3.27}$$

which implies

$$|g(u, y)| \leq 2(R + M)^2 D\sigma^n \leq 8R^2 D\sigma^n := c_R. \tag{3.28}$$

Hence $|g(u, y) - \mu(g)| \leq 2c_R$.

It follows from the Schwarz inequality that

$$\begin{aligned} |g(u, y)|^2 &= \left| \int_X \Phi\left(\frac{x}{\sigma}, \frac{u}{\sigma}\right) (f(u) - f_{\mathcal{H},\sigma,x}(u))(f(u) + f_{\mathcal{H},\sigma,x}(u) - 2y) d\rho_X(x) \right|^2 \\ &\leq \int_X \Phi\left(\frac{x}{\sigma}, \frac{u}{\sigma}\right) (f(u) - f_{\mathcal{H},\sigma,x}(u))^2 (2R + 2M)^2 d\rho_X(x) \\ &\quad \times \int_X \Phi\left(\frac{x}{\sigma}, \frac{u}{\sigma}\right) d\rho_X(x). \end{aligned} \tag{3.29}$$

By (3.27),

$$\mu(g^2) \leq 16R^2 D\sigma^n \int_X \left(\int_X \Phi\left(\frac{x}{\sigma}, \frac{u}{\sigma}\right) (f(u) - f_{\mathcal{H},\sigma,x}(u))^2 d\rho_X(u) \right) d\rho_X(x). \tag{3.30}$$

It has been proved in [9] that

$$\begin{aligned} &\int_X \Phi\left(\frac{x}{\sigma}, \frac{u}{\sigma}\right) (f(u) - f_{\mathcal{H},\sigma,x}(u))^2 d\rho_X(u) \\ &= \int_Z \Phi\left(\frac{x}{\sigma}, \frac{u}{\sigma}\right) [(f(u) - y)^2 - (f_{\mathcal{H},\sigma,x}(u) - y)^2] d\rho(u, y). \end{aligned} \tag{3.31}$$

Substituting (3.31) into (3.30),

$$\begin{aligned} \mu(g^2) &\leq 16R^2 D\sigma^n \int_X \left(\int_Z \Phi\left(\frac{x}{\sigma}, \frac{u}{\sigma}\right) [(f(u) - y)^2 - (f_{\mathcal{H},\sigma,x}(u) - y)^2] d\rho(u, y) \right) d\rho_X(x) \\ &= 16R^2 D\sigma^n \int_Z \left(\int_X \Phi\left(\frac{x}{\sigma}, \frac{u}{\sigma}\right) [(f(u) - y)^2 - (f_{\mathcal{H},\sigma,x}(u) - y)^2] d\rho_X(x) \right) d\rho(u, y) \\ &= 16R^2 D\sigma^n \mu(g). \end{aligned} \tag{3.32}$$

Using Proposition 3.4 with $\alpha = \frac{1}{4}$ and $\mathcal{G} = \mathcal{G}_R$, we know that

$$\begin{aligned} &P \left\{ \sup_{f \in B_R} \frac{\int_X [(\mathcal{E}_x(f) - \mathcal{E}_x(f_{\mathcal{H},\sigma,x})) - (\mathcal{E}_{z,x}(f) - \mathcal{E}_{z,x}(f_{\mathcal{H},\sigma,x}))] d\rho_X(x)}{\sqrt{\varepsilon + \int_X (\mathcal{E}_x(f) - \mathcal{E}_x(f_{\mathcal{H},\sigma,x})) d\rho_X(x)}} \geq \sqrt{\varepsilon} \right\} \\ &\leq (1 + 4e^{-2\bar{\alpha}}) \mathcal{N}\left(\mathcal{G}_R, \frac{\varepsilon}{4}\right) \exp\left\{-\frac{3m^{(\alpha)}\varepsilon}{2048R^2 D\sigma^n}\right\}. \end{aligned} \tag{3.33}$$

Since for any $g_1, g_2 \in \mathcal{G}_R$,

$$\begin{aligned}
 |g_1(u, y) - g_2(u, y)| &= \left| \int_X \Phi\left(\frac{x}{\sigma}, \frac{u}{\sigma}\right) ((f_1(u) - y)^2 - (f_2(u) - y)^2) d\rho_X(x) \right| \\
 &\leq \left| \int_X \Phi\left(\frac{x}{\sigma}, \frac{u}{\sigma}\right) (f_1(u) - f_2(u)) \right. \\
 &\quad \left. \times (f_1(u) + f_2(u) - 2y) d\rho_X(x) \right| \\
 &\leq 4RD\sigma^n |f_1(u) - f_2(u)|,
 \end{aligned} \tag{3.34}$$

then we have

$$\begin{aligned}
 \mathcal{N}\left(\mathcal{G}_R, \frac{\varepsilon}{4}\right) &\leq \mathcal{N}\left(B_R, \frac{\varepsilon}{16RD\sigma^n}\right) \\
 &= \mathcal{N}\left(B_1, \frac{\varepsilon}{16R^2D\sigma^n}\right).
 \end{aligned} \tag{3.35}$$

It follows from (3.33) that

$$\begin{aligned}
 P\left\{ \sup_{f \in B_R} \frac{\int_X [(\mathcal{E}_x(f) - \mathcal{E}_x(f_{\mathcal{H}, \sigma, x})) - (\mathcal{E}_{z,x}(f) - \mathcal{E}_{z,x}(f_{\mathcal{H}, \sigma, x}))] d\rho_X(x)}{\sqrt{\varepsilon + \int_X (\mathcal{E}_x(f) - \mathcal{E}_x(f_{\mathcal{H}, \sigma, x})) d\rho_X(x)}} \leq \sqrt{\varepsilon} \right\} \\
 \geq 1 - (1 + 4e^{-2\bar{\alpha}}) \mathcal{N}\left(B_1, \frac{\varepsilon}{16R^2D\sigma^n}\right) \exp\left\{-\frac{3m^{(\alpha)}\varepsilon}{2048R^2D\sigma^n}\right\}.
 \end{aligned} \tag{3.36}$$

We set the term $(1 + 4e^{-2\bar{\alpha}}) \mathcal{N}(B_1, \frac{\varepsilon}{16R^2D\sigma^n}) \exp\{-\frac{3m^{(\alpha)}\varepsilon}{2048R^2D\sigma^n}\}$ of the above inequality to $\delta/2$. We need to invoke the lemma proved by the same method of Proposition 4.3 in [21].

Lemma 3.2 *Let $\eta^*(m^{(\alpha)}, \delta)$ be the smallest positive solution of the following inequality in η :*

$$(1 + 4e^{-2\bar{\alpha}}) \mathcal{N}(B_1, \eta) \exp\left\{-\frac{3m^{(\alpha)}\eta}{128}\right\} \leq \delta.$$

If $\log \mathcal{N}(B_1, \eta) \leq c_p(\eta)^{-p}$, for some $p \in (0, 2)$, $c_p > 0$ and all $\eta > 0$, then with confidence at least $1 - \delta$, we have

$$\eta^*(m^{(\alpha)}, \delta) \leq \max\left\{ \frac{256}{3m^{(\alpha)}} \log \frac{1 + 4e^{-2\bar{\alpha}}}{\delta}, \left(\frac{256c_p}{3m^{(\alpha)}}\right)^{1/(1+p)} \right\}. \tag{3.37}$$

Then we return to the proof of Proposition 3.5.

It follows from Theorem 5.3 in [27] that

$$\mathcal{N}\left(B_1, \frac{\varepsilon}{16R^2D\sigma^n}\right) \begin{cases} \leq \left(\frac{32R^2D\sigma^n}{\varepsilon} + 1\right)^{\bar{d}}, & \text{for } 0 < \varepsilon < 16R^2D\sigma^n; \\ = 1, & \text{for } \varepsilon \geq 16R^2D\sigma^n. \end{cases} \tag{3.38}$$

When $0 < \varepsilon < 16R^2D\sigma^n$,

$$\begin{aligned} \log \mathcal{N}\left(B_1, \frac{\varepsilon}{16R^2D\sigma^n}\right) &\leq \tilde{d} \log\left(\frac{32R^2D\sigma^n}{\varepsilon} + 1\right) \\ &\leq \frac{\tilde{d}}{p} \left(\frac{32R^2D\sigma^n}{\varepsilon}\right)^p \\ &= \frac{\tilde{d}}{p} (32R^2D\sigma^n)^p \varepsilon^{-p}, \quad p > 0. \end{aligned} \tag{3.39}$$

When $\varepsilon \geq 16R^2D\sigma^n$, we have

$$\log \mathcal{N}\left(\mathcal{G}_R, \frac{\varepsilon}{4}\right) \leq 0. \tag{3.40}$$

Hence we conclude that

$$\log \mathcal{N}\left(B_1, \frac{\varepsilon}{16R^2D\sigma^n}\right) \leq \frac{2^p \tilde{d}}{p} \left(\frac{\varepsilon}{16R^2D\sigma^n}\right)^{-p} := c_p \eta^{-p}. \tag{3.41}$$

This, together with (3.37), implies that, for

$$m^{(\alpha)} \geq \frac{256}{3} c_p^{-\frac{1}{p}} \left(\log \frac{2 + 8e^{-2\bar{\alpha}}}{\delta}\right)^{1+\frac{1}{p}}, \tag{3.42}$$

we obtain

$$\varepsilon \leq 16R^2D\sigma^n \left(\frac{256c_p}{3m^{(\alpha)}}\right)^{1/(1+p)}. \tag{3.43}$$

Combining (3.43) with (3.36), with confidence $1 - \delta/2$, we have

$$\begin{aligned} &\int_X [(\mathcal{E}_x(f) - \mathcal{E}_x(f_{\mathcal{H},\sigma,x})) - (\mathcal{E}_{z,x}(f) - \mathcal{E}_{z,x}(f_{\mathcal{H},\sigma,x}))] d\rho_X(x) \\ &\leq \sqrt{\varepsilon} \sqrt{\varepsilon + \int_X (\mathcal{E}_x(f) - \mathcal{E}_x(f_{\mathcal{H},\sigma,x})) d\rho_X(x)} \\ &\leq \varepsilon + \frac{1}{2} \int_X (\mathcal{E}_x(f) - \mathcal{E}_x(f_{\mathcal{H},\sigma,x})) d\rho_X(x) \\ &\leq 16R^2D\sigma^n \left(\frac{256c_p}{3m^{(\alpha)}}\right)^{1/(1+p)} + \frac{1}{2} \int_X (\mathcal{E}_x(f) - \mathcal{E}_x(f_{\mathcal{H},\sigma,x})) d\rho_X(x). \end{aligned} \tag{3.44}$$

Finally, setting $f = f_{z,\sigma,x}$ in the above inequality, we derive the desired result. □

4 Proofs of the main results

In this subsection, we provide the proofs of Theorem 2.1 and Theorem 2.2. We firstly prove Theorem 2.1.

Proof If we take $\sigma = (m^{(\alpha)})^{-\gamma}$, $\gamma > 0$, then we have

$$R = \max\{C_{\mathcal{H},\rho_X}, C'_{\mathcal{H},\rho_X}, M\} m^{\gamma(\zeta + \frac{\zeta}{2} \max\{1, \zeta\})}. \tag{4.1}$$

It is readily seen that (2.3) implies (3.1). Then Proposition 3.5 holds true. We thus obtain, with confidence $1 - \delta$,

$$\int_X (\mathcal{E}_x(f_{\mathcal{Z},\sigma,x}) - \mathcal{E}_x(f_{\mathcal{H},\sigma,x})) d\rho_X(x) \leq 32R^2 D\sigma^n \left(\frac{256c_p}{3m^{(\alpha)}}\right)^{1/(1+p)}. \tag{4.2}$$

Therefore from (2.8) we obtain

$$\begin{aligned} \|f_{\mathcal{Z}} - f_{\mathcal{H}}\|_{\rho_X}^2 &\leq \widetilde{C}_{\mathcal{H}} \sigma^{-2\zeta - \tau \max\{\zeta, 1\}} \times 32R^2 D\sigma^n \left(\frac{256c_p}{3m^{(\alpha)}}\right)^{1/(1+p)} \\ &\leq \widetilde{C}(m^{(\alpha)})^{4\gamma(\zeta + \frac{\tau}{2} \max\{\zeta, 1\}) - n\gamma - \frac{1}{p+1}}. \end{aligned} \tag{4.3}$$

This proves Theorem 2.1. □

Next, we prove Theorem 2.2.

Proof Let $\gamma = \varepsilon/[4\zeta + 2 \max\{\tau, \tau\zeta\}] > 0$ and $p = \frac{\varepsilon}{1-\varepsilon}$. Therefore we have

$$\gamma \tau \max\{1, \zeta\} < \varepsilon < 1/4 \tag{4.4}$$

and

$$(m^{(\alpha)})^{1-\gamma \tau \max\{1, \zeta\}} \geq (m^{(\alpha)})^{1-\varepsilon} \geq m^{\frac{1}{2}}. \tag{4.5}$$

It follows from (2.5) that

$$\begin{aligned} (m^{(\alpha)})^{\frac{1}{2}} &\geq \left(A_{\tau, \zeta} + \frac{256}{3} c_p^{-\frac{1}{p}}\right) \left(1 + \frac{1}{4\zeta + 2 \max\{\tau, \tau\zeta\}}\right) \\ &\quad \times \left((\log(2 + 8e^{-2\bar{\alpha}})/\delta)^{1+\frac{1}{p}} + \log m^{(\alpha)}\right) \\ &\geq \left(A_{\tau, \zeta} + \frac{256}{3} c_p^{-\frac{1}{p}}\right) \left((\log(2 + 8e^{-2\bar{\alpha}})/\delta)^{1+\frac{1}{p}} + \gamma \log m^{(\alpha)}\right), \end{aligned} \tag{4.6}$$

which implies that condition (2.3) of Theorem 2.1 holds true, we thus obtain, with confidence $1 - \delta$,

$$\|f_{\mathcal{Z}} - f_{\mathcal{H}}\|_{\rho_X} \leq C_1 (m^{(\alpha)})^{\varepsilon - \frac{1}{2}}. \tag{4.7}$$

This proves Theorem 2.2. □

Funding

This work was supported by the National Natural Science Foundation of China (grant no. 11671213) and the Fundamental Research Funds for the Central Universities.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

Both authors contributed equally to the writing of this paper. Both authors read and approved the final manuscript.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 9 March 2018 Accepted: 21 July 2018 Published online: 31 July 2018

References

- Cerny, M., Antoch, J., Hladik, M.: On the possibilistic approach to linear regression models involving uncertain, indeterminate or interval data. *Inf. Sci.* **244**(7), 26–47 (2013)
- Fasshauer, G.E.: Toward approximate moving least squares approximation with irregularly spaced centers. *Comput. Methods Appl. Mech. Eng.* **193**(12–14), 1231–1243 (2004)
- Komargodski, Z., Levin, D.: Hermite type moving-least-squares approximations. *Comput. Math. Appl.* **51**(8), 1223–1232 (2006)
- McLain, D.H.: Drawing contours from arbitrary data points. *Comput. J.* **17**(17), 318–324 (1974)
- Savitzky, A., Golay, M.J.E.: Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **36**(8), 1627–1639 (1964)
- Wand, M.P., Jones, M.C.: *Kernel Smoothing*. Chapman & Hall, London (1995)
- Shepard, D.: A two-dimensional interpolation function for irregularly-spaced data. In: *ACM National Conference*, pp. 517–524 (1968)
- Wang, H.Y., Xiang, D.H., Zhou, D.X.: Moving least-square method in learning theory. *J. Approx. Theory* **162**(3), 599–614 (2010)
- Wang, H.Y.: Concentration estimates for the moving least-square method in learning theory. *J. Approx. Theory* **163**(9), 1125–1133 (2011)
- He, F.C., Chen, H., Li, L.Q.: Statistical analysis of the moving least-squares method with unbounded sampling. *Inf. Sci.* **268**(1), 370–380 (2014)
- Tong, H.Z., Wu, Q.: Learning performance of regularized moving least square regression. *J. Comput. Appl. Math.* **325**, 42–55 (2017)
- Guo, Q., Ye, P.X.: Error analysis of the moving least-squares method with non-identical sampling. *Int. J. Comput. Math.*, 1–15 (2018). <https://doi.org/10.1080/00207160.2018.1469748>
- Smale, S., Zhou, D.X.: Online learning with Markov sampling. *Anal. Appl.* **7**(1), 87–113 (2009)
- Pan, Z.W., Xiao, Q.W.: Least-square regularized regression with non-iid sampling. *J. Stat. Plan. Inference* **139**(10), 3579–3587 (2009)
- Guo, Z.C., Shi, L.: Classification with non-i.i.d. sampling. *Math. Comput. Model.* **54**(5), 1347–1364 (2011)
- Guo, Q., Ye, P.X.: Coefficient-based regularized regression with dependent and unbounded sampling. *Int. J. Wavelets Multiresolut. Inf. Process.* **14**(5), 1–14 (2016)
- Sun, H.W., Guo, Q.: Coefficient regularized regression with non-iid sampling. *Int. J. Comput. Math.* **88**(15), 3113–3124 (2011)
- Chu, X.R., Sun, H.W.: Half supervised coefficient regularization for regression learning with unbounded sampling. *Int. J. Comput. Math.* **90**(7), 1321–1333 (2013)
- Steinwart, I., Hush, D., Scovel, C.: Learning from dependent observations. *J. Multivar. Anal.* **100**(1), 175–194 (2009)
- Billingsley, P.: Convergence of probability measures. *Appl. Stat.* **159**(1–2), 1–59 (1968)
- Wu, Q., Ying, Y.M., Zhou, D.X.: Learning rates of least-square regularized regression. *Found. Comput. Math.* **6**(2), 171–192 (2006)
- Lv, S.G., Feng, Y.L.: Semi-supervised learning with the help of Parzen windows. *J. Math. Anal. Appl.* **386**(1), 205–212 (2012)
- Sun, H.W., Wu, Q.: Regularized least square regression with dependent samples. *Adv. Comput. Math.* **32**(2), 175–189 (2010)
- Chu, X.R., Sun, H.W.: Regularized least square regression with unbounded and dependent sampling. *Abstr. Appl. Anal.* **2013**, Article ID 139318 (2013)
- Guo, Q., Ye, P.X., Cai, B.L.: Convergence rate for l^p -coefficient regularized regression with non-i.i.d. sampling. *IEEE Access* **6**, 18804–18813 (2018)
- Parthasarathy, K.R.: Convergence of probability measures. *Technometrics* **12**(1), 171–172 (1968)
- Cucker, F., Zhou, D.X.: *Learning Theory: An Approximation Theory Viewpoint*. Cambridge University Press, Cambridge (2007)
- Modha, D.S., Masry, E.: Minimum complexity regression estimation with weakly dependent observations. *IEEE Trans. Inf. Theory* **42**(6), 2133–2145 (1996)